

## MULTIMEDIA GLOSSES AND THEIR EFFECT ON L2 TEXT COMPREHENSION AND VOCABULARY LEARNING

**Iñigo Yanguas**

**San Diego State University**

The present study investigates the effects that different types of multimedia glosses, namely textual, pictorial, and textual + pictorial, have on text comprehension and vocabulary learning when the goal is exclusively comprehension of a computerized text. This study is based on the theoretical framework of attention, which maintains that attention is critical in the acquisition process of an L2 (Robinson, 1995; Schmidt, 1995, 2001; Tomlin and Villa, 1994). Ninety-four participants read a text under one of four gloss conditions while asked to think aloud. This study investigated whether any of the conditions promoted noticing and whether this noticing led to better comprehension of the text and learning of the target vocabulary words. Reading comprehension, recognition, and production measures were utilized in a pre-post test design. Results of quantitative and qualitative analyses of the data gathered showed first that all multimedia gloss groups noticed and recognized significantly more of the target words than the control group. Second, no significant differences were found among any of the groups in production of the target vocabulary items. Finally, regarding comprehension, results showed that the combination gloss group significantly outperformed all other groups. These results confirm that the multimedia glosses under investigation have a different effect on comprehension and vocabulary learning respectively.

### INTRODUCTION

During the past twenty years, a considerable number of studies have investigated the effectiveness of marginal glosses under different premises (Bowles, 2004; Chun & Plass, 1996; Hulstijn, Hollander, & Greidanus, 1996; Jones, 2004; Jones & Plass, 2002; Lomicka, 1998; Nagata, 1999) and various theoretical frameworks support their use as a valid tool for second language acquisition (SLA) (Hulstijn & Laufer, 2001; Mayer, 2001, 2002, 2005b; Schmidt, 1990 and elsewhere). Glosses, usually placed at the margin of the text, can be textual, visual, both textual and visual, or auditory. The present study addresses the first three types within the theoretical framework utilized in Bowles (2004), so that its results can have a wider scope of interpretation.

Researchers have mainly investigated two issues regarding annotations in general: on the one hand, whether they do in fact aid students' overall comprehension of the text, and which exact type of gloss is more beneficial; on the other hand, studies have tried to determine if these types of annotations promote incidental vocabulary learning. The most common measures in these studies are "off-line" recall comprehension protocols and post reading vocabulary tests (e.g., Al-Seghayer, 2001). However, given that off-line protocols may be dependent on memory and, therefore, may not fully reflect learners' comprehension processes, "online measures" currently employed in SLA research conducted within an attentional framework (e.g., Leow, 1997, 2000; Rosa & Leow, 2004; Rosa & O'Neill, 1999) will be employed in this study to investigate what type of available gloss (textual, pictorial, or a combination) better aids L2 (second language) students to comprehend a written passage and acquire more vocabulary words incidentally.

## LITERATURE REVIEW

### Glosses as an aid for overall comprehension of the text

Glosses act as substitutes for the dictionary. Glosses, however, do not interrupt the reading process as much, since the definition is easily available in the text. Traditionally, they have been used to promote comprehension of the text and incidental vocabulary learning (Bowles, 2004). In general, when comparing marginal glosses, whether multimedia or traditional, with the absence of any type of gloss, glosses have been shown to be of help to the student in the comprehension of a written text (Bowles, 2004; Davis, 1989; Jacobs, DuFon, & Hong, 1994; Lomicka, 1998). These studies, however, utilized very different research designs so that interpreting their results as a whole is not an easy task.

For example, Davis (1989) investigated whether marginal glosses improved comprehension of a literary text read in a foreign language by intermediate-level college students of French. Results showed that both presenting the students with a vocabulary guide before reading the text and providing glosses when reading the text helped students recall significantly more of the reading passage than those with no help.

Lomicka (1998) investigated whether glossing aided L2 comprehension of a written text, and whether glossing hindered fluency in the L2. She was also interested in the relationship between the number of glosses consulted and the level of comprehension achieved, as well as inferences created. She stated that the most common measure in the literature to assess the influence of glosses had been recall protocols, and she argued that this type of measure was not a reliable method for representing underlying processes or comprehension, for that matter. She drew on the use of online protocols from the L1 literature to assess the creation of causal inferences to argue that this kind of measurement was a more valid approach. Metalinguistic think-alouds allowed her to check whether the students made causal inferences from the glosses and created a situation model. Twelve participants enrolled in a second semester French course at an American university took part in the pilot study and were randomly assigned to one of the three conditions: a control group, a group with access to traditional glosses, and a group with access to multimedia extended glosses. Participants in all conditions thought aloud to check for inferences and comments made about the structure of the text or the glosses. Statistical analyses indicated no significant differences in the number of inferences created by each group. The author concluded, however, that the data gathered from the think-aloud protocols indicated that multimedia annotations seemed to have a positive effect on comprehension of a written text and the construction of a situation model. She argued that her results were not significant due to the small number of participants and that future research was needed to further investigate these issues.

Bowles (2004) also employed think-aloud protocols and compared computerized glosses versus traditional glosses in vocabulary acquisition and text comprehension measures. However, her motivation to do so came from a very different theoretical background, namely, Schmidt's (1990) noticing hypothesis, which is also used in the present study. Schmidt claimed that learners have to "notice" the form in question before it can be processed further. To measure noticing, Bowles used online protocols to establish what type of gloss induced learners to "notice" more words, and to investigate whether this potential increase in noticing had an impact on text comprehension and vocabulary acquisition. A final sample of 50 participants enrolled in first year college-level Spanish courses was randomly assigned to one of three groups: a control group, a traditional (textual) gloss group, and a multimedia gloss group. Participants' performances on a comprehension task, and immediate and delayed recognition and written production tasks were submitted to quantitative and qualitative analyses. The results indicated that both experimental groups, multimedia and traditional gloss, had an advantage over the control in number of noticed words, text comprehension, and acquisition of target vocabulary. These results therefore seemed to support the use of multimedia or traditional glosses, but they showed no significant difference between them.

At the other end of the spectrum, Jacobs, DuFon, and Hong (1994) found that glossing did not significantly affect recall in general. These researchers explored the effect of vocabulary glossing on recall and vocabulary learning as well as the attitudes and preferences for the two types of gloss investigated (English and Spanish glosses). The main findings in this study were that glossing did not significantly affect recall, although there was a trend that favored students who had access to glosses. However, post hoc analysis of the scores on the recall measure showed that those students with higher proficiency recalled more if they had had access to a glossed word. In the translation task, those who had glosses outperformed those with no access to glosses. Regarding vocabulary learning, superior scores for those students who were presented with glosses disappeared after four weeks.

Although Jacobs et al.'s findings would seem to contradict what was reported by the studies reviewed above, the fact is that the post-reading assessment tasks were not similar: participants in this study were asked to write everything they could remember about the text and write translations of the target words. It could therefore be argued that these types of measures do not assess the degree of comprehension or acquisition, since they rely heavily on short-term memory. Another possible explanation for the contradictory results could be provided looking at the differences in proficiency level across studies. It seems then feasible that annotations have a different impact on learners with different proficiency in the L2.

After reviewing these studies, it could be stated that glosses have a general beneficial effect on L2 reading comprehension. However, different research designs should lead us to interpret these results with caution. To avoid this problem, the present study utilizes Bowles' (2004) research design and theoretical framework but expands the types of glosses investigated so that the study's results can be easily interpreted in a wider context.

### **The influence of multimedia annotations on vocabulary learning when the goal is reading comprehension**

Incidental vocabulary learning has been defined as the picking up of new words during listening or reading activities when the general goal of the activity is comprehension rather than specifically learning those new words (Hulstijn et al. 1996). These authors set out to investigate whether providing glosses and repeating the same words in a text could offset the small gains generally found from incidental vocabulary learning. The researchers concluded that frequency of occurrence had a greater beneficial effect on incidental vocabulary learning both when the gloss had been accessed and when the word had been looked up in the dictionary than when no meaning of the word was available. In addition, marginal glosses better promoted incidental vocabulary learning, given that in some cases students did not use the dictionary. When dictionaries were indeed used, the results were as good as with glosses. However, half of the words that had been glossed, looked up, or appeared between one and three times within the text were forgotten. In summary, this study supported the positive role of annotations on incidental vocabulary learning.

Drawing from the L2 input enhancement literature and L1 reading literature, Watanabe (1997) explored the effects of text modifications (appositives and single and multiple-choice marginal glosses) and task on incidental L2 vocabulary learning when reading. Results confirmed that students who had access to either glossing format outperformed those participants who had not had access to glosses, in three unexpected vocabulary posttests. In addition, no significant results were found for the scores of those participants who could choose the type of gloss and those who were provided with the translation.

Finally, Rott and Williams (2003) utilized think-aloud protocols when qualitatively exploring the effect of multiple-choice glosses and periodic second language text reconstruction on lexical acquisition. These authors investigated learners' processing strategies and how access to a number of glosses influenced inferencing strategies and form-meaning mapping. They also investigated the effects of number of occurrences of target words on noticing and form-meaning mappings as well as the impact of an output

activity on these variables. Twelve fifth-semester L1 English learners of German at college-level formed the final sample of participants in this study. These participants were randomly assigned to one of two conditions: + gloss, - gloss. The reading passage included four target words, for which participants were pretested, and seven distractors. The participants had to retell in German each of the four sections of the text immediately after reading it. To measure immediate vocabulary gain, two immediate vocabulary posttests were administered. In addition, in order to assess L2 readers' processing behavior, a concurrent think-aloud procedure was used.

In general, the results of this study revealed a difference in reading strategies and vocabulary test scores for both groups. In particular, those students who had access to glosses looked for concrete meanings and strong form-meaning mapping; students with no access to glosses aimed at overall comprehension of the text; if students re-encountered the target words their form-meaning connections would be strengthened.

Rott and Williams' study, however, has some limitations that should be kept in mind when interpreting its results. First, only 12 students took part in this experiment, seven participants in one group and five in the other; second and more importantly, it lacks an operationalization of "noticing" and depth of processing. The theoretical framework on which the study was developed was Hulstijn and Laufer's (2001) Involvement Load Hypothesis. Hulstijn and Laufer's basic contention was that "the retention of unfamiliar words is, generally, conditional upon the degree of involvement in processing these words" (p. 545). "On line" think-aloud protocols are one tool that researchers have made use of to gain insight into cognitive processes and noticing. In this study, however it was implied that it was the gloss and the kind of activity that promoted noticing and certain depth of processing but lacked the operationalization of those variables. Consequently, there is no objective way to address the degree of processing or noticing.

In summary, as seen in the research reviewed, the claims concerning vocabulary learning have to be considered carefully, since these studies do not attempt to capture any internal cognitive process that might lead to learning. Moreover, they are based on a small number of respondents. Following Bowles (2004), the present study seeks to aid interpretation by utilizing Schmidt's (1990) noticing hypothesis that posits that forms have to be noticed to be further processed.

### **What type of gloss is more effective for reading comprehension and vocabulary learning?**

Several types of glossing when reading for comprehension have been addressed in the literature and could be subsumed under textual and visual glossing and a mixture of the two. Texts, pictures, or dynamic videos placed at the margin of texts when the students click on the hyperlinked word help the reader understand the meaning of difficult words in the text. The first two types and their combination will be addressed in the present study.

In a widely cited study, Chun and Plass (1996) investigated, on the one hand, incidental vocabulary learning when the goal was reading comprehension, and on the other, the effect of visual information for vocabulary learning and students' look-up behavior. Their results were threefold: firstly, incidental vocabulary acquisition rates of 25% in production tasks and 77% in recognition tasks were observed; secondly, significantly higher scores were found for words annotated with text and picture than those glossed with text only or text + video; thirdly, a significant weak to moderate positive correlation between using a certain annotation type and using this same type for the retrieval of words was found.

Al-Seghayer (2001) investigated what type of multimedia gloss, dynamic video or still picture, was more effective in aiding vocabulary acquisition. Results of the analyses carried out indicated that learners presented with text + video had scores that were significantly higher than the text + picture condition. In addition, results also showed that the text only condition was significantly more effective than the text + video condition.

Kost, Foss, and Lenzi (1999) also measured how different types of glosses influenced incidental vocabulary growth in a non-multimedia environment. Results showed that those participants who had

access to both textual and pictorial glosses outperformed those under the textual and pictorial gloss conditions in the recognition of target words on both short-term memory and retention tests.

The authors argued that these results were due to the different degrees of cognitive effort needed to process: “the mapping of pictures onto the mental model provides a stronger bond than the mapping of words due to the different representations of their information” (p. 94).

Plass, Chun, Mayer, and Leutner (1998) investigated the effects of different types of glosses according to the students’ preferred mode on text comprehension and learning of new words. The results of this study provided justification for the generative theory of multimedia learning (Mayer, 1997), an earlier version of Mayer’s cognitive theory of multimedia learning (Mayer, 2005b). Participants performed better on the posttests when both visual and textual information were selected, moderate when only one mode was selected, and worse when neither was selected. In addition, participants comprehended the text better when they could choose the gloss in their preferred mode.

Based on work carried in the field of psychology, Mayer proposed that it is through two channels that human beings represent and manipulate knowledge: a visual-pictorial and an auditory-verbal channel (Mayer, 1997, 2001, 2002, 2005b). Therefore, the use of textual and pictorial glosses would enter the cognitive system through those two channels. He argued that “meaningful learning occurs when learners engage in active processing within the channels, including selecting relevant words and pictures, organizing them into coherent pictorial and verbal models, and integrating them with each other and appropriate prior knowledge” (Mayer, 2002, p. 60). His edited volume on multimedia learning (Mayer, 2005a) includes additional developments to the cognitive theory of multimedia learning (CTML) that he developed. This theory is based on three general cognitive science principles: the dual-channel assumption, the limited capacity assumption, and the active processing assumption. According to these principles, humans process information through different channels according to its nature, have a limited processing capacity that makes the allocation of cognitive resources key in the learning process, and have an active role in processing of information by paying attention, organizing, and integrating new information (Mayer, 2005b). In addition, Plass and Jones (2005) discussed the applications that this theory has for SLA. These authors propose a model based on the CTML and on an interactionist model of SLA in which attention must be paid in order for learning to take place. In their view, multimedia contexts facilitate the provision of meaningful input, foster interaction, and provide opportunities to elicit output in the L2.

To sum up, the literature on different types of multimedia glosses appears to indicate that the combination of textual and visual glosses has a more beneficial effect on comprehension and vocabulary learning than either type in isolation. As mentioned above, however, any generalization about vocabulary learning should be closely considered, since those studies reviewed above utilize measures that involve memory skills to a great degree and lack an operationalization of any level of cognitive awareness. In addition, the different theoretical backgrounds from which the above studies originate and the different research designs make difficult any general conclusion as to what types of glosses are more beneficial for the L2 learner in reading a computerized text.

Based on Schmidt’s (1990) noticing hypothesis, the present study operationalized and measured attention through think-aloud protocols following previous studies (e.g., Bowles, 2004; Leow, 1997, 2000; Rosa & Leow, 2004; Rosa & O’Neill, 1999). Critics might argue, however, that by introducing concurrent verbal protocols as a tool to investigate internal cognitive processes as they occur these very same processes could be being altered. This is no place to discuss this in depth, it should suffice to say that this issue of reactivity is still to be settled, since there are some studies using different types of tasks that show reactivity (e.g., Jourdenais, 2001) and several others that do not (e.g., Leow, 2000, where an elaborated discussion of the benefits of online/concurrent versus offline data elicitation procedures can be found). As

Pintrich and Schunk (1996) pointed out “there is no one, best type of self-report measurement, the choice of the instrument must match the purpose for the assessment and the research problem” (p. 19).

This study is then a conceptual replication of Bowles (2004), with additional independent variables such as the types of multimedia glosses discussed above, so that its results can be interpreted in a broader context. By using Schmidt’s noticing hypothesis and operationalizing attention under this theoretical framework, this study addresses one of the main research design flaws of previous studies in this strand of research that based their conclusions on recall measures. Participants read a text under four conditions (no gloss, text gloss, picture gloss, and text + picture gloss) while asked to think aloud. These different conditions were used to assess whether any of them promoted noticing and whether this noticing led to a better comprehension of the text and vocabulary learning of the target words. Specifically, answers to the following research questions were sought:

## RESEARCH QUESTIONS

1. Do readers in any of the groups report noticing targeted vocabulary words significantly more than readers in any of the other groups?
2. Does exposure to any type of gloss have a significant effect on L2 readers’ learning of target vocabulary words as measured through a) recognition or b) production tasks? If so, will this effect be maintained over a period of three weeks?
3. Does exposure to any type of gloss have a significant effect on L2 readers’ comprehension of a computerized reading passage, as measured by a comprehension task?

## METHOD

### Participants

The original pool of participants consisted of nine sections of fourth semester college-level Spanish, a total of 133 participants. As in Bowles (2004), participants were students at a small private university located on the Northeastern Seaboard. Participants were fulfilling their last semester of foreign language requirement. A final pool of 94 participants remained after participants were eliminated for the following reasons: 1) scoring more than 6 out of 21 on the pretests (a chance score; 4 participants); 2) indicating outside exposure during the study on the debriefing questionnaire (administered at the end of the study; 3 participants); 3) failing to produce intelligible think-aloud protocols (8 participants); 4) failing to attend all sessions (9 participants); and 5) failing to record answers (15 participants).

### Materials

#### *Computerized reading text*

An Internet-based passage from an online newspaper was utilized. To minimize the risk of content familiarity affecting results a current piece of news was used: Greenpeace en la Antártica. This was an authentic text from the online version of the Spanish newspaper *El Mundo*; the length of the text was 543 words (see [Appendix A](#); the whole text can be read when the different types of glosses are shown). Macromedia Flash MX 2004 was used to design the text, which was then uploaded online so that participants could access it. The text was divided into four paragraphs, each one on a different page; participants could navigate back and forth between the pages (see [Appendix A](#)).

*Piloting of the experimental text:* Nineteen randomly selected participants underlined all the unknown words in the text. Only words underlined by more than half of the participants were selected to be glossed.

*Glossing:* As a result of the pilot test 21 words were glossed. In the experimental conditions, the words were hyperlinked. When the participants clicked on them a box appeared above the word with a definition

in English (textual gloss group), a picture (pictorial gloss group), or a combination (textual + pictorial gloss group). Participants in the control group were exposed to the text with no glosses.

### **Assessment tasks**

The 21 annotated words in the reading text were the focus of the recognition and production tasks. The production tasks (pre-, immediate posttest, and delayed posttest) included the same 40 items (21 target words plus 19 distractors) in scrambled orders. The recognition tasks (pre-, immediate posttest, and delayed posttest) included only the 21 target words. Given the multiple-choice nature of this task, distractors were included in the multiple-choice options (see [Appendix A](#)). Following Bowles (2004), no delayed comprehension posttest was included in the research design of this study; therefore, recall of the text after three weeks could not be investigated.

#### ***Pretest /Posttest production tasks*** (see sample questions in [Appendix B](#))

The same 40-item production task was utilized in the pretest and immediate and delayed posttests. Participants were given a list of words in English (21 target words and 19 distractors) and asked to write their equivalent in Spanish. All the production tasks differed in the order in which the items appeared. Participants performed the production task first so that the multiple-choice recognition task would not provide additional exposure to the target words.

#### ***Pretest / Posttest recognition tasks*** (see sample questions in [Appendix B](#))

The same 21-item multiple-choice recognition task was utilized in the pretest and immediate and delayed posttests. The target words in Spanish were followed by four possible equivalents in English. As in the production tasks, all three recognition tasks differed in the order of items. Given the multiple-choice nature of the task, distractors were included in the options given.

#### ***Comprehension task*** (see sample questions in [Appendix B](#))

The comprehension task was administered only in the immediate posttest. It consisted of 11 multiple-choice comprehension questions in English. Nine of these questions were specifically focused around the glossed words; the remaining two were of a global nature. Participants were allowed to refer back to an unglossed version of the text to help them answer the comprehension questions if they needed to. As can be seen in the sample questions in [Appendix B](#), the multiple-choice comprehension questions used demanded that learners understand the specific meaning of the words glossed in the text, since some of the choices are very close in meaning. The two questions dealing with overall comprehension of the text aimed at measuring whether participants had achieved a clear understanding of events happening in the text where no glosses had been used.

### **Procedure**

Participants were randomly assigned to one of the four gloss groups. Two weeks prior to exposure, participants took a paper version of the production and recognition pretests to assess their knowledge of the target words. On the day of exposure, participants in the three groups were given instructions to complete the experimental treatment, including instructions and practice on how to think aloud while performing the task. In addition, a second version of the production and recognition tasks was administered as immediate posttest. This test also included the comprehension task described above. Three weeks after exposure, a delayed posttest with scrambled items was administered. To control for outside exposure to the target words, a debriefing questionnaire was administered after the delayed posttest. In this questionnaire, participants were asked whether they had had any significant amount of exposure to the Spanish vocabulary or topic used in this investigation along the duration of the study. After 39 participants were excluded for one or more of the reasons given above, the final number of participants per group was the following: control, N=23; picture & text gloss, N=25; picture gloss, N=26; text gloss, N=20.

*Coding think-aloud data:* Two raters coded the data according to the following criteria (Bowles, 2004): any comments made about the target word, any gloss read aloud, or any mention of any hyperlinked word clicked were used as operationalizations of “noticing.”

## Results

Prior to analyzing participants’ scores, reliability coefficients were computed on the vocabulary and comprehension tasks using Cronbach’s alpha. For the recognition and comprehension tasks coefficients were relatively high (recognition task: .68; comprehension task: .72). Nevertheless, the reliability coefficient for the production task was low (.29). This could be attributed to the restricted distribution of scores (Bowles, 2004).

Numbers of instances of reported noticing for each group out of a total possible of 21 (means: control =4.08, SD= 3.18; picture & text= 16.36, SD=3.03; picture= 15.57, SD= 3.77; text= 13.70, SD=3.45) were submitted to a one-way ANOVA ([research question 1](#)). Results are shown in [Table 1](#) below:

Table 1. Noticing ANOVA for Group

Source	<i>df</i>	Sum of squares	Mean squares	<i>F</i>	<i>p</i>
Group	3	2274.846	758.282	66.378	.0001
Total	93	3302.979			

As can be seen, results showed a main effect for Group. A post hoc Scheffé test (see [Table 2](#)) revealed that this main effect was due to a significant difference ( $p = .0001$ ) between the control group (mean=4.08) on the one hand, and the experimental groups on the other (mean= 13.70, 15.57, and 16.36 respectively). In other words, this indicates that participants who received one of the types of glosses noticed significantly more vocabulary words than those who received an unglossed text. However, there was no significance in the amount of noticing between the gloss groups. Following Cohen (1998), large effect sizes were found:  $d > .80$ .

Table 2. Reported Noticing: Post Hoc Scheffé’s for Type of Gloss

Type	N	Subset for alpha = 0.05	
		1	2
Control	23	4.08	
Text Gloss	20		13.70
Picture Gloss	26		15.57
Picture & Text Gloss	25		16.36
Sig.		1.000	.073

Qualitative analyses of the think-aloud protocols showed that most participants in the gloss groups noticed target vocabulary words at a low level of awareness. In other words, participants presented with glosses did not attempt any kind of deeper processing of the vocabulary items glossed. Along the lines of Bowles (2004), the vast majority of instances of noticing were simply acknowledgements of the meaning of the word. For example, one participant in the picture gloss group stated after accessing the gloss that showed somebody hiding: “Umm, this should be to hide.” Another participant after accessing the picture glosses for *ballena* (whale) and *rorcual aliblanco* (minke whale) made clear that her only concern was meaning: “Ok, these two pictures show whales...I guess *ballena* is whale and this other longer word is some type of whale....” In the text gloss group, there were also some clear examples of participants being



interested in just getting the meaning of the words with no indication of any lexical process: (participant looking at the definition gloss given for *embistió*) "...attacked the boat...whaaat?...oh OK got it." Further on, this same participant struggled to understand part of a paragraph even if she acknowledged the meaning of the gloss given for *popa* (stern): "...after the action of...[mumbling]...this kind of boat...ok...the chief of the exhibition...[mumbling]...ok...he said that ecologists...wait...la popa...oh the stern...realized the...the...[pause] [laughing] the Japanese navy...huh...I don't know what's going on...hold on..." It seems clear from this excerpt that this participant was mainly worried about the general meaning of the sentence giving little importance to the words glossed other than acknowledging their meaning. In the mix gloss group, learners seemed to also focus on meaning. For example, one participant in this condition made it quite clear: "well, that does it, this little note is definitely clearer than that picture...to hide...the picture is good though."

For [research question 2a](#), recognition scores for every group were submitted to a repeated measures ANOVA in which Group was entered as the between-subject variable and Time as the within-subject variable (pretest means and SD: control =5.04, SD= .92; picture & text= 5.16, SD=1.24; picture= 5.38, SD= .75; text= 5.25, SD= .63; posttest means and SD: control =8.17, SD= 1.55; picture & text= 15.57, SD=3.77; picture= 14.16, SD = 2.59; text= 13.55, SD=3.03; delayed test means and SD: control =6.43, SD= 1.44; picture & text= 10.32, SD =1.81; picture= 10.57, SD= 2.26; text= 9.95, SD=1.93). Results are shown in [Table 3](#) below:

Table 3. Recognition ANOVA for Group and Time

Source	df	Sum of squares	Mean squares	F	p
Group	3	549.972	183.324	30.309	.0001
Time	1	786.300	786.300	367.547	.0001
Time x Group	3	117.419	39.140	18.295	.0001

This analysis yielded significant main effects for Group, Time, and the interaction of both, as can be seen in [Table 3](#) above, also visually displayed on [Figure 1](#) below. The significant interactions appears to be due to the significant increases experienced by the experimental groups when compared to the control group. A post hoc Scheffé test (see [Table 4](#) below) revealed that this main effect ( $p = .0001$ ) was due to a significant difference between the control group (mean=6.55) on the one hand and the experimental groups on the other (mean=9.58, 9.85, and 9.88 respectively). More than half of the variance in the recognition scores is attributable to Group (Eta squared= .503). However, there was no significant difference between the three glossing conditions.

Table 4. Recognition Scores: Post Hoc Scheffé's for Type of Gloss

(I) TYPE	(J) TYPE	Mean difference (I-J)	Std. Error	Sig.
Control	Picture & Text Gloss	-3.32*	.4102	.000
	Picture Gloss	-3.30*	.4064	.000
	Text Gloss	-3.03*	.4341	.000

\*The mean difference is significant at 0.05

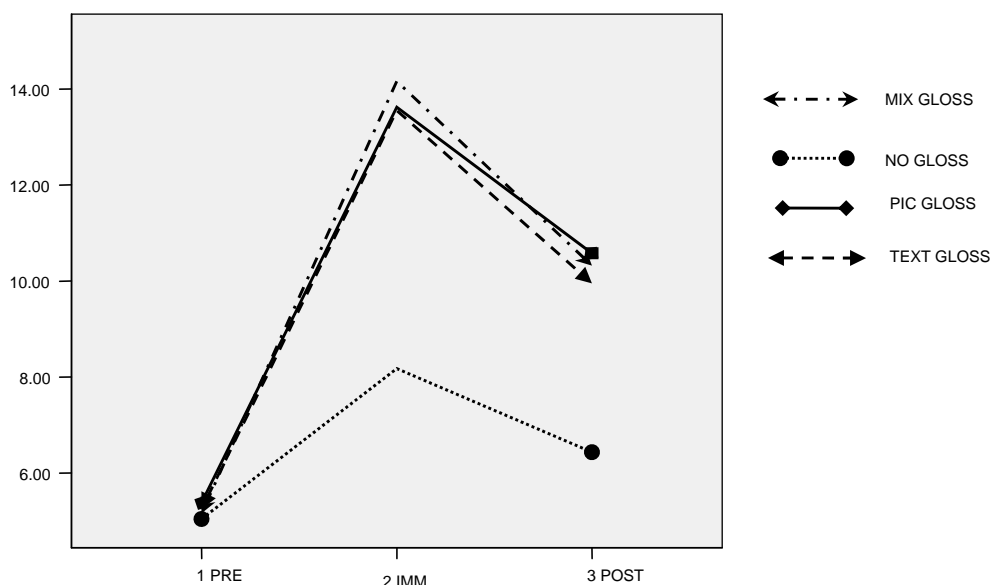


Figure 1. Mean recognition scores across time

To ascertain whether the decrease on the delayed posttest impacted the significant main effect for Time shown in the above analysis, further repeated measures ANOVA analyses were conducted for each group with Time as the within-subjects factor and word recognition scores as dependent variables. Results of these analyses showed a significant main effect for Time for every group (control:  $F(2, 44)=75.806$ ;  $p: 0001$ ; Eta squared= .775; picture & text:  $F(2, 48)=133.807$ ;  $p: 0001$ ; Eta squared= .848; picture:  $F(2, 50)=154.381$ ;  $p: 0001$ ; Eta squared= .861; text:  $F(2, 38)=117.133$ ;  $p: 0001$ ; Eta squared= .860). As shown in the Eta squared results, more than 77% of the variance on the recognition scores for every group can be attributable to Time. In addition, to further investigate this main effect for Time, a series of paired t-tests were performed so that possible significant growth or loss across test sessions could be assessed. Results of these t-tests showed significant increases in recognition scores from pre- to posttests for all groups (control:  $t=-10.78$ ,  $p= .0001$ ; picture & text  $t=-15.28$ ,  $p= .0001$ ; picture  $t=-14.19$ ,  $p= .0001$ ; text  $t=-12.25$ ,  $p= .0001$ ). Furthermore, significant decreases were shown in these scores from post- to delayed posttests for all groups (control:  $t=10.29$ ,  $p= .0001$ ; picture & text  $t=6.83$ ,  $p= .0001$ ; picture  $t=8.12$ ,  $p= .0001$ ; text  $t=7.06$ ,  $p= .0001$ ). Finally, these paired t-test analyses showed a significant increase from pre- to delayed posttests for all groups (control:  $t=-4.85$ ,  $p= .0001$ ; picture & text  $t=-10.27$ ,  $p= .0001$ ; picture  $t=-11.69$ ,  $p= .0001$ ; text  $t=-11.43$ ,  $p= .0001$ ).

Results summarized above suggest a positive answer to [research question 2a](#), that is, exposure to the multimedia glosses utilized in this study appeared to have a differential effect on learners' ability to recognize the target vocabulary words, and these differential effects tended to be maintained over a period of three weeks.

For [research question 2b](#), production scores were submitted to a repeated measures ANOVA (pretest means and SD: control = .91, SD= .79; picture & text= 1.00, SD= .70; picture= 1.00, SD= .80; text= 1.10, SD= .55; posttest means and SD: control =1.21, SD= .73; picture & text= 2.04, SD=1.24; picture= 2.00, SD= 1.38; text= 2.00, SD=1.25; delayed test means and SD: control = .95, SD= .82; picture & text= 1.44, SD= .82; picture= 1.42, SD= .75; text= 1.70, SD= .92). Results yielded only a significant main effect for Time, with neither significant main effect for Group nor a significant interaction between Time and Group ([Table 5](#)). [Figure 2](#) below visually represents these results.

Table 5. Production ANOVA for Group and Time

Source	<i>df</i>	Sum of squares	Mean squares	<i>F</i>	<i>p</i>
Group	3	12.980	4.327	2.118	.103
Time	1	6.600	6.600	25.681	.0001
Time x Group	3	1.853	.618	2.403	.073

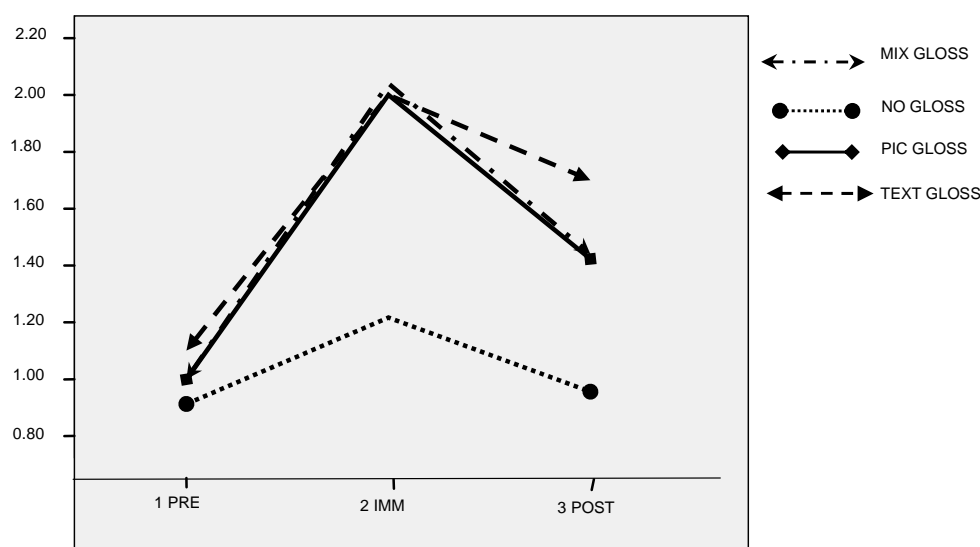


Figure 2. Mean production scores across time

As in the recognition scores above, further repeated measures ANOVA analyses were computed with the production scores as dependent variables and Time as within-subject factor for every group. These analyses were conducted to assess whether the decrease in the delayed posttest had an impact on the significant increase found in the analyses above. Results showed a significant main effect for Time for every group (control:  $F(2, 44)=5.059$ ;  $p=.011$ ; Eta squared=.187; picture & text:  $F(2, 48)=19.192$ ;  $p=.0001$ ; Eta squared=.444; picture:  $F(2, 50)=16.743$ ;  $p=.0001$ ; Eta squared=.401; text:  $F(2, 38)=15.545$ ;  $p=.0001$ ; Eta squared=.450). Results showed a main effect for Time for every group so that the significant increase found in the previous analysis seems not to be impacted by the decrease in the delayed posttests. However, the amount of variance on the production scores that can be attributed to Time is rather small (less than 45% for all groups) as shown by the Eta squared results. As with the recognition scores above, further paired t-test analyses were carried out to assess significant score growth or loss across tests. Results of these t-tests showed significant increases in production scores from pre- to posttests for all groups (control:  $t=-3.10$ ,  $p=.005$ ; picture & text  $t=-5.31$ ,  $p=.0001$ ; picture  $t=-5.43$ ,  $p=.0001$ ; text:  $t=-4.41$ ,  $p=.0001$ ). Furthermore, significant decreases were shown in these scores from post- to delayed posttests for all groups (control:  $t=2.78$ ,  $p=.011$ ; picture & text  $t=3.92$ ,  $p=.001$ ; picture  $t=3.26$ ,  $p=.003$ ; text:  $t=2.34$ ,  $p=.030$ ). Finally, these paired t-test analyses showed a significant increase from pre- to delayed posttests for all groups except the control group (control:  $t=-3.71$ ,  $p=.009$ ; picture & text  $t=-2.86$ ,  $p=.009$ ; picture  $t=-2.26$ ,  $p=.013$ ; text:  $t=-3.94$ ,  $p=.001$ ).

These results suggest that exposure to multimedia glosses in this study did not appear to have a differential effect on learners' ability to produce the target vocabulary items. Results showed, however, a

differential effect of Time on their production ability, which was maintained after a period of three weeks for all groups except the control group.

Finally, a one-way ANOVA was computed on comprehension scores to determine whether type of gloss condition had a significant effect on text comprehension. (means: control =4.13, SD= 1.21; picture & text= 8.08, SD=1.70; picture= 6.73, SD= 1.51; text= 6.25, SD=1.37). Results showed a significant main effect for Group (Table 6). A post hoc Scheffé test revealed that this main effect ( $p = .0001$ ) was due to a significant difference between the control group (mean=4.13) on the one hand and the experimental groups on the other (mean=6.25, 6.73, and 8.08 respectively). In addition, this test showed that the combination gloss group (mean=8.08) significantly outperformed all others (combination gloss group and control group,  $p = .0001$ ; combination gloss group and text gloss group,  $p = .001$ ; combination gloss group and picture gloss group,  $p = .017$ ). Following Cohen (1998), small effect sizes were found:  $d < .50$ .

Table 6. Comprehension ANOVA for Group

Source	df	Sum of squares	Mean squares	F	p
Group	3	192.101	64.034	29.506	.0001
Total	93	378.415			

Table 7. Comprehension Scores: Post Hoc Scheffé's for Type of Gloss

Type	N	Subset for alpha = 0.05		
		1	2	3
Control	23	4.1304		
Text Gloss	20		6.2500	
Picture Gloss	26		6.7308	
Picture & Text Gloss	25			8.0800
Sig.		1.000	.744	1.000

## DISCUSSION

In response to [research question one](#), results of the present study showed that participants exposed to multimedia glosses reported noticing the target vocabulary words significantly more than participants in the control group. However, no significant difference was found among any of the gloss groups.

These results are consistent with Bowles (2004) who found that both traditional and computerized glosses aided in noticing significantly more target forms when compared to a control group but found no difference between gloss groups. She claimed that those findings are not surprising, given that glosses are designed to draw the reader's attention and help in comprehending unknown vocabulary items. Along the same lines, in the present study, no significant differences between gloss groups were found. Cognitively, it could be argued that glosses that present the information visually or textually might cause readers to process information through different channels and, therefore, differences in processing might be feasible (Mayer, 1997 and elsewhere). However, as in previous studies (Bowles, 2004; Lomicka, 1998; Rott & Williams, 2003) analyses of the think-aloud protocols allowed an insight into participants' cognitive processes as they interacted with the glosses, which provide a possible explanation of these results: all multimedia annotations in this study aided respondents in noticing the target vocabulary items at a low level of awareness. In general, participants in all three gloss groups focused on meaning when accessing the gloss and went on reading without evidence of further processing: "Ok, this means to attack." They did stop when they felt that they were not following what they were reading, which is further proof that

their main concern was understanding the text: “Ummm, so what is it saying? ...that these guys on a boat tried to...” On the contrary, around 70 % of participants in the control group overlooked all of the unglossed target items. Their reading behavior, as shown in the think-aloud protocols, was similar to the other groups’ and focused on comprehension. In this group, two major trends could be distinguished: those that either read for themselves or read aloud but stopped whenever a comprehension problem arose, and those that read the whole text without stopping. For the most part, however, they stopped only when overall comprehension was at stake, not at any given unknown word: “...let’s see, I think this means that there are some big companies involved...” Some participants in this group, however, did mention the words they did not understand while looking for overall meaning of the sentence. Some of these words were the target words, some were not: “I guess it’s probably that the Arctic Sunrise was virtually stopped at that time...[mumbling] el barco Esperanza and the Arctic Sunrise are boats that protect the ocean I guess... I don’t know caso mismo.... *caso omiso* means [pause] I’m not exactly clear what *además* is....”

In relation to [research question two](#), which investigated the effect of type of gloss on production and recognition tasks and whether this effect was maintained over a period of three weeks, no significant effect was found in the production task. When interpreting these results, however, the low internal consistency of this task has to be borne in mind. As far as word recognition is concerned, the three gloss groups significantly outperformed the control group even though no significant difference between experimental groups was found. These results are not surprising and coincide with previous findings that have shown production of target vocabulary items not to be affected by the appearance of annotations (Bowles, 2004; Chun & Plass, 1996; Kost et al., 1999).

As argued above, participants seemed to focus on text comprehension, leaving aside word morphology or any type of lexical association (Bowles, 2004; Lomicka, 1998) which would be indicative of processes that might lead to vocabulary acquisition. The lack of these processes might explain why no significant differences were found in the production task and qualitative analyses of think-alouds support this statement; glosses seemed to help participants make sense of the general meaning of the sentence but not make any type of lexical association that might signal deep processing of the word. This behavior by participants in the gloss groups seemed to be enough to perform significantly better than the control group in the recognition tests but not in the production tests. These results contradict findings in previous studies (Hulstijn et al., 1996; Watanabe, 1997) in which they argued that the appearance of glosses promotes vocabulary learning. Both studies, however, measured vocabulary learning through recall and post-vocabulary tests only, so that any deeper interpretation of those results is not available. Furthermore, Hulstijn et al. (1996) included in their research design frequency of occurrence, which turned out to be a key variable in their final results. Frequency of occurrence is therefore a variable that could be investigated in this context to assess its impact on production of L2 vocabulary.

With respect to the issue of time, a significant main effect was shown for both recognition and production tasks. Learners in all experimental conditions for both production and recognition tasks significantly increased their scores from the pretest to the immediate posttest but also showed a significant loss from the immediate to the delayed posttest three weeks later. However, this negative effect is counterbalanced by the fact that participants still showed a significant gain in scores from the pretest to the delayed posttest. It could be said, therefore, that the differential effects shown for gloss type on production and comprehension measures tend not to be maintained over a period of three weeks. A possible interpretation of these results could be in the way participants approached the processing of target vocabulary words as shown in the analysis of the think aloud protocols. Participants did not seem to attempt to make any lexical analysis or connection signaling deeper processing of the words that could have potentially led to longer lasting learning.

Finally, concerning the [third research question](#) in which the effect of different types of glosses is investigated in relation to text comprehension, all groups exposed to glosses outperformed the control group. In addition, the combination gloss group outperformed all others. In other words, comprehension

of the text was better aided by visual and textual information as provided by combination glosses. This outcome is very interesting and clearly contrasts with the above results regarding production and recognition, in which no difference between the different gloss groups was found. This seems to indicate that multimedia glosses have a different effect on vocabulary acquisition and text comprehension respectively. It seems that glosses do not offer enough support to fully learn the target vocabulary item but significantly aid in comprehending the text. Once again, analyses of think-aloud protocols provide further explanation of the results achieved. Some participants in the picture gloss group seemed to think that the pictures were not very helpful, even distracting: "...I don't get the pictures...they are annoying...." Some other participants had trouble comprehending some of the pictures: "...*esconder*?...what is this girl doing?...this picture is weird...." In the mix gloss group, where students get a text gloss along the picture, participants did not have to interpret what the picture meant and most of them read aloud the definition given while making a side comment about the picture: "[reading]...*embistió*...[reading the text gloss and then commenting on the picture]...rush forward, attack...ok...wow!...indeed!...this bullfighter is being attacked...." The combination gloss group had access to both pictures and text, and most participants took advantage of it. The appearance of both text and picture seemed to have a cognitive impact on participants' reading behavior: one participant, for example, recorded: [laughing] "nice picture... [reading] to deceive." However, it eventually only impacted reading comprehension, as shown by the quantitative results.

These cases would help explaining differences between the mix gloss group and the picture gloss group, since it seems that being able to access pictures was not very helpful for some of the participants. Pictures are less explicit than definitions and somehow open to interpretation, and some of the participants found it harder to get a suitable meaning for the picture accessed in the gloss. Nevertheless, if we accepted this interpretation of the results we would still need to explain why participants in the mix gloss group significantly outscored participants in the text gloss group. Learners in this group accessed glosses, read them, had no trouble comprehending them and went on reading the text. As participants in other groups, they only stopped when they tried to make sense of longer chunks of the text. One participant, for example, stated while reading with the help of text glosses: "the only reason I can figure this out is because of the pictures (meaning the text glosses)...I wish they had more of it...I don't understand though [reading, pause]...huh...[reading gloss]...*edible fish or shellfish from the sea*...[pause]...[translating]...Greenpeace wants...these actions...announced that...so I'm guessing...huh...Greenpeace has like huh...." It seems plausible, therefore, that providing learners with both picture and text glosses so that they could process information through different channels, as proposed by Mayer (2001, 2005), allowed participants to better comprehend the text.

These final results are consistent with findings in several studies that have found combination glosses more beneficial for incidental vocabulary learning and text comprehension (Al-Seghayer, 2001; Kost et al., 1999; Plass et al., 1998). Recall measures were not utilized in this study to assess learning and text comprehension; instead, concurrent verbal protocols were used in addition to recognition, production, and comprehension measures so that a clearer picture of the cognitive processes taking place could be gained. The present results would then provide support for Mayer's cognitive theory of multimedia learning (Mayer, 2001; 2005b) and its SLA conceptualization (Plass & Jones, 2005) that accounts for differences channels to process textual and pictorial input. As shown by the present results, it is when both channels are engaged that better comprehension of reading passage occurs.

## CONCLUSION

At least three important implications can be derived from results of this study. First, it has been shown that glosses have a different impact on vocabulary learning and text comprehension respectively. It seems that a combination gloss is more beneficial for text comprehension only. Second, by being a conceptual replication of Bowles (2004), the present study strengthens this strand of research so that the results from

both studies can be interpreted in a wider context. Both of these studies operationalized and measured “noticing”—an internal process argued to be of great importance in foreign language learning (Schmidt, 1990). By doing this, this study addresses one of the major limitations of previous studies, namely, the failing to operationalize that attention was indeed paid to the target forms before its effect was statistically analyzed. Third, think-aloud protocols have turned out to be very useful tools to investigate how readers make use of multimedia glosses and have made possible a clearer interpretation of the results than through quantitative analyses alone.

## **LIMITATIONS AND FURTHER INVESTIGATION**

There are three main limitations to the present study. First, the low reliability coefficients of the production task items that make us interpret the results on production with caution. Second, the lack of measurement for the time spent on task by every group. Finally, some participants’ comments seem to suggest that some of the pictures were difficult to understand, which could have affected the results. To counter this limitation, what glosses were accessed how many times and for how long could have been controlled. The provision of research design components to account for these limitations would have provided this strand of research with very valuable information and make clearer the interpretation of the present results. They should be taken into consideration in future studies.

Moreover, the following aspects could be added to this study’s research design to widen the scope of interpretation. First, video glosses have been investigated in the literature and could be included in an experimental design of this nature. By being more dynamic and explicit, they may have a different effect on foreign language readers. Second, proficiency level is a variable that should be investigated in the future, given that the effect of glosses could be different across proficiency levels. Third, the different effect of L1 and L2 glosses could be also investigated, since they have been shown to impact long-term retention of vocabulary items in a different way (Yoshii, 2006). Fourth, frequency of occurrence has also been previously investigated in relation to incidental vocabulary learning and it would be a very interesting line of research along with multimedia glosses. Last but not least, Mayer’s cognitive theory of multimedia learning provides an ideal theoretical framework to investigate multimedia learning and the cognitive processes involved in L2 learning. Future studies should be designed to assess its different components.

---

## **ACKNOWLEDGMENTS**

I would like to thank all the colleagues and friends that have contributed in different ways to the completion of this article. Also, I am grateful to the reviewers for their insightful comments that have definitely improved this study.

---

## **ABOUT THE AUTHORS**

Iñigo Yanguas is an assistant professor of Spanish and director of the Spanish Language Program at San Diego State University. His research interests include different aspects of SLA, CALL and Spanish as a heritage language in the US. Currently, he is investigating the place of motivation in the language acquisition process and vocabulary learning in Oral CMC.

E-mail: [iyanguas@mail.sdsu.edu](mailto:iyanguas@mail.sdsu.edu)

---



## APPENDIXES

### Appendix A. Sample Multimedia Glosses

#### Picture gloss

glosaspicturesBb.swf

2 of 6



Varias lanchas de Greenpeace se interpusieron entre el arpón del ballenero japonés Kyo Maru y una ballena (un rorcual aliblanco) en el Santuario Ballenero de la Antártida para impedir su caza. Esta es la tercera acción de la organización ecologista en los últimos días en la zona, a donde ha desplazado dos de sus mayores barcos para tratar de impedir las 'capturas científicas' que cada año realiza Japón. Una de las lanchas neumáticas del barco MY Esperanza trató de interponerse entre el rorcual y el arpón del ballenero, exhibiendo una pancarta con el lema "Sponsored by Gorton's" ("patrocinado por Gorton's") sobre la ballena cazada para denunciar el vínculo entre la caza de estos animales y esta empresa de marisco estadounidense. Greenpeace quiere con estas acciones denunciar que bajo la denominación de "caza científica" de la flota japonesa se esconden claros intereses comerciales.

previous

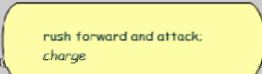
THINK ALOUD!!

next

#### Text gloss

glosastextuales.swf

3 of 6



Según ha denunciado la propia organización, después de esta acción el buque-factoría japonés Nisshin Maru embistió el barco Artic Sunrise. El jefe de la expedición de Greenpeace, Shane Rattenbury, dijo a la cadena australiana ABC que 25 ecologistas viajaban a bordo de la embarcación y que ninguno resultó herido a causa de la embestida por la popa que realizó la nave japonesa, de 7.575 toneladas y una tripulación de 115 personas. "No había otro navío en la zona y no había razón alguna para que nos enfilara. El Artic Sunrise estaba virtualmente parado en ese momento", comentó Rattenbury al citado medio. Ésta es la tercera acción de la organización en el Santuario Ballenero de la Antártida desde que comenzó la campaña de caza de ballenas, el pasado 20 de noviembre, fecha en la que la organización ecologista comenzó su expedición en defensa de los océanos con el barco Esperanza y el Arctic Sunrise.

previous

THINK ALOUD!!

next



**Combination gloss**

MIXglosses.swf

4 of 6



To cause something or something to stop

En el último... las protestas... hecho caso omiso de los... tos de la Comisión Ballenera Internacional (CBI) para que detengan su programa de "caza científica de ballenas", sino que ha decidido unilateralmente aumentar su cuota de captura a más del doble de ejemplares de rorcuales aliblanco hasta un total de 935. Además, ha añadido 10 ejemplares del rorcual común, una especie amenazada. Esta ballena es la segunda criatura más grande del planeta, después de la ballena azul.

previous THINK ALOUD !! next

**No gloss**

singlossBb.swf

5 of 6

"Es inaceptable que continúen con una caza de ballenas innecesaria y completamente injustificada. No pueden seguir engañando con el argumento de la caza científica cuando todos sabemos que estas ballenas ya se están troceando y empaquetando para ser vendidas", ha declarado Sebastián Losada, responsable de Océanos de Greenpeace. Greenpeace está utilizando todos los medios de que dispone para acabar con la caza de ballenas a corto plazo y conseguir que ésta sea la última vez que los balleneros ataquen el Santuario. Esto incluye seguir la pista a quienes están beneficiándose económicamente de las capturas, como en este caso la empresa Gorton's. La campaña en defensa de las ballenas en el Santuario Ballenero de la Antártida es la primera etapa de una nueva y ambiciosa campaña de Greenpeace, "En defensa de los océanos". Todo el año 2006, el Esperanza será la principal plataforma de Greenpeace para exponer la necesidad de crear una red de reservas mundial.

FIN

previous THINK ALOUD !! next

**Appendix B. Sample Questions**

---

**Word Recognition Test**

Choose the English equivalent of the Spanish words below. **KEEP THINKING ALOUD!**

- 1) ballena  
a) dolphin                      b) orca                      c) minke whale                      d) whale
- 2) engañando  
a) threatening                      b) halting                      c) hiding                      d) deceiving
- 

**Word Production Test**

Directions: In the blank next to each English word, type the Spanish equivalent. Do not worry about accents. **KEEP THINKING ALOUD!**

1. boat                      \_\_\_\_\_
2. hunt                      \_\_\_\_\_
3. whale                      \_\_\_\_\_
- 

**Comprehension Test**

Now **answer the questions** ON THE FOLLOWING TWO PAGES based on the information provided in the passage. **KEEP THINKING ALOUD.**

1. What does the article say about the Esperanza?
- a) That it will be Greenpeace's main platform.
  - b) That Greenpeace will shut it down.
  - c) That Greenpeace will track it down.
  - d) That Greenpeace will aid it.
3. Paragraph 3: What happened between the Nisshin Maru and the Arctic Sunrise?
- a) The Nisshin Maru attacked the Arctic Sunrise.
  - b) The Nisshin Maru avoided the Arctic Sunrise.
  - c) The Nisshin Maru aided the Arctic Sunrise's crew.
  - d) The Nisshin Maru towed the Arctic Sunrise.

## REFERENCES

- Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202-232. Retrieved February 7, 2008, from <http://lt.msu.edu/vol5num1/alseghayer/default.html>.
- Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania*, 87(3), 541-552.
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80(2), 183-198.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates
- Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal*, 73(1), 41-48.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327-339.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539-558.
- Jacobs, G. M., DuFon, P., & Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, 17(1), 19-28.
- Jones, L. (2004). Testing L2 vocabulary recognition and recall using pictorial and written test items. *Language Learning & Technology*, 8(3), 122-143. Retrieved February 7, 2008, from <http://lt.msu.edu/vol8num3/jones/default.html>.
- Jones, L., & Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546-561.
- Jourdenais, R. (Ed.). (2001). *Cognition, instruction, and protocol analysis*. Cambridge: Cambridge University Press.
- Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, 32(1), 89-113.
- Leow, R. P. (1997). Attention, awareness and foreign language behavior. *Language Learning*, 47, 467-505.
- Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware vs. unaware learners. *Studies in Second Language Acquisition*, 22, 557-584.
- Lomicka, L. L. (1998). "To Gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 41-50. Retrieved February 7, 2008, from <http://lt.msu.edu/vol1num2/article2/default.html>.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1-19.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. In D. F. Halpern & M. D. Hakel, (Eds.), *Applying the science of learning to university teaching and beyond* (pp. 55-72). San Francisco: Jossey-Bass.

- Mayer, R.E. (Ed.). (2005a). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.
- Mayer, R. E. (2005b). Cognitive theory of multimedia learning. In Mayer, R.E. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.
- Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, 32(4), 469-479.
- Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research, and applications*. Englewood Cliffs, NJ: Merrill Prentice-Hall.
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90, 25-36.
- Plass, J., & Jones, L. (2005). Multimedia learning in second language acquisition. In Mayer, R.E. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467-488). New York: Cambridge University Press.
- Robinson, P. (1995). Attention, memory and the “noticing” hypothesis. *Language Learning*, 45(2), 283-331.
- Rosa, E., & Leow, R. P. (2004). Awareness, different language conditions, and second language development. *Applied Psycholinguistics*, 25, 269-292.
- Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. Another piece to the puzzle. *Studies in Second Language Acquisition*, 21(4), 511-556.
- Rott, S., & Williams, J. (2003). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language*, 15(1). Retrieved February 15, 2005, from <http://nflrc.hawaii.edu/rfl/April2003/rott/rott.pdf>.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R. (1995). Consciousness and foreign language learning. In R. Schmidt, (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). University of Hawai'i at Manoa: Second Language Teaching and Curriculum Center.
- Schmidt, R. (2001). Attention. In P. Robinson, (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge Applied Linguistics Series. Cambridge University Press.
- Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16, 183-203.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19(3), 287-307.
- Yoshii, M. (2006). L1 and L2 Glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85-101. Retrieved February 7, 2008, from <http://llt.msu.edu/vol10num3/yoshii/default.html>.